Acoustic correlates of plosive voicing in Madurese

Acoustic correlates of plosive voicing in Madurese

Misnadin[1] and James Kirby[2]

[1)] *Department of English, Universitas Trunojoyo Madura, Jl. Raya Telang, Kecamatan Kamal, Bangkalan, Madura 69162 Indonesia*[a]

[2] *School of Philosophy, Psychology, and Language Science, University of Edinburgh, Dugald Stuart Building, 3 Charles Street, Edinburgh EH8 9AD U.K.*[b]

Madurese, a Malayo-Polynesian language of Indonesia, is of interest both areally and typologically: it is described as having a three-way laryngeal contrast between voiced, voiceless unaspirated, and voiceless aspirated plosives, along with a strict phonotactic restriction on consonant voicing-vowel height sequences. We present an acoustic analysis of Madurese consonants and vowels obtained from recordings of fifteen speakers, to assess whether its voiced and aspirated plosives might share acoustic properties indicative of a shared articulatory gesture. Although we find that voiced and voiceless aspirated plosives in word-initial position pattern together in terms of several spectral balance measures, these are most likely due to the following vowel quality, rather than aspects of a shared laryngeal configuration. Conversely, the voiceless (aspirated and unaspirated) plosives share multiple acoustic properties, including $F0$ trajectories and overlapping voicing lag time distributions, suggesting that they share a glottal aperture target. We discuss the implications of these findings for the typology of laryngeal contrasts and the historical evolution of the Madurese consonant-vowel co-occurrence restriction.

[a] misnadin@trunojoyo.ac.id

[b] j.kirby@ed.ac.uk

I. INTRODUCTION

A. Background

Madurese is a Western Malayo-Polynesian language spoken primarily on the island of Madura and a number of regions in East Java, Indonesia. The language may be roughly divided into three mutually intelligible dialect regions, Western, Central, and Eastern (Kiliaan, 1897; Soegianto *et al.*, 1986; Stevens, 1968). Of these, Eastern Madurese is considered as the standard dialect and is taught at elementary and junior high schools across Madura and the regencies along the northern coast of East Java. Madurese is spoken by an estimated 8 to 15 million speakers, making it the fourth largest language spoken in Indonesia after Indonesian, Javanese and Sundanese (Davies, 2010).

While there exist several treatments of Madurese phonology, morphology, and syntax (Davies, 2010; Kiliaan, 1897; Stevens, 1968), comparatively little attention has been focused on the phonetic structures of this language. The only published acoustic analyses are those of Cohn and colleagues (Cohn, 1993a,b; Cohn and Ham, 1999; Cohn and Lockwood, 1994), which are based on the speech of just two native speakers. But Madurese displays several areally and typologically unusual properties that deserve further detailed study, both for what they can reveal about the language itself, as well as for what they can teach us about the typology of laryngeal contrast more generally.

First, Madurese is described as having a three-way laryngeal contrast between voiced, voiceless unaspirated, and voiceless aspirated plosives at five places of articulation (Table I). This is unexpected given that its geographically neighbouring and genetically related languages uniformly have a two-way contrast between either unaspirated and prevoiced plosives, as in Indonesian (Adisasmito-Smith, 2004) or Sundanese (Kulikov, 2010), or between so-called 'stiff' and 'slack' voice qualities as in Javanese (Fagan, 1988; Thurgood, 2004). If Madurese truly makes a three-way laryngeal contrast, it is unusual: languages contrasting prevoiced, voiceless unaspirated, and voiceless aspirated plosives – individually all quite commonly attested – appear to be comparatively rare, and tend more often than not to be tonal (Kirby, 2018a). For example, none of the languages considered in the survey of Cho and Ladefoged (1999) are of this type. Moreover, the phonetic properties of plosives in three-way systems are often underspecified, and may reflect an incomplete understanding of laryngeal articulations and their acoustic consequences (Seyfarth and Garellek, 2018). The status of the Madurese 'voiceless aspirated' stops is a case in point: several orthographies represent this series as *bh, dh, ḍh, jh, gh*, and the phonetic transcriptions in some current dictionaries (e.g. Pawitra, 2009) transcribe these as voiced aspirates rather than voiceless aspirates, but Cohn and Lockwood (1994) did not find any evidence of voicing during the closure phase of these segments. A more detailed understanding of such systems could enhance our understanding of laryngeal typology. Second, the distribution of plosives in Madurese is highly restricted. Phonetically, the language distinguishes eight vowel qualities [i ɛ a ə ɨ ɔ ɤ u] (Cohn, 1993b; Misnadin and Kirby, 2018; Stevens, 1968). The 'high vowels' [i ɤ u ɨ] are always preceded by a voiced or voiceless aspirated plosive (hereafter /D/ and /TH/, respectively), while the 'non-high vowels' [ɛ a ɔ ə] occur elsewhere: word-initially, following a voiceless unaspirated plosive (hereafter /T/), or (with some exceptions) a sonorant, /s/,

Table I. Consonant system of Madurese, after (Misnadin and Kirby, 2018). Consonants in parenthesis () are canonically restricted to loanwords.

	Bilabial	Dental/ Alveolar	Retroflex	Palatal	Velar	Glottal
Plosive	p	t	ʈ	c	k	ʔ
	p^h	t^h	$ʈ^h$	c^h	k^h	
	b	d	ɖ	ɟ	ɡ	
Nasal	m	n		ɳ	ŋ	
Fricative	(f)	s				(h)
Lateral		l				
Trill		r				
Glide	w			j		

or /ʔ/. The eight surface vowel qualities of Madurese can thus be analysed in terms of four high/non-high pairs (Table II).

While this distribution might suggest that [p^h t^h $ʈ^h$ c^h k^h] are simply allophones of /p t ʈ c k/ which surface before high vowels, morphophonological evidence clearly favors an analysis with three levels of voicing and four vowel pairs. The primary evidence supporting this account is that while phonetic vowel height can always be predicted given the identity of a preceding consonant, the converse is not always the case. For example, when the actor

Table II. CV co-occurrence restriction in Madurese. For each pair, the first example shows the voiceless unaspirated plosive /T/ plus non-high vowel, and the second and third examples show the aspirated /TH/ and voiced /D/ plosives plus high vowels.

ɛ ∼ i	pɛɾak	‘happy’	a ∼ ɤ	padɤ	‘same’
	p^{h}iʈak	‘bird’		p^{h}ɤʈɛ	‘profit’
	bisa	‘able’		bɤca	‘read’
ɔ ∼ u	pɔtɛ	‘white’	ə ∼ ɨ	pəsːɛ	‘money’
	p^{h}uta	‘giant’		p^{h}ɨsːɛt	‘scratched’
	buta	‘blind’		bɨsːɛ	‘iron’

voice morpheme /N/ is prefixed to a stem, it surfaces with a place of articulation homorganic to the following consonant, but is also always followed by a non-high vowel: /N/ + [bɤbɤ] ‘low’ → [mabɤ], /N/ + [patɛ] ‘die‘ → [matɛ], but /N/ + [p^{h}ɤkta] ‘bring’ → [makta]. If ‘bring’ is underlyingly /pɤkta/, one must explain why the actor voice prefix lowers the vowel in ‘bring’ but not in ‘low’. In addition, the high vowels [i ɨ ɤ u] never occur in absolute word-initial position. This distributional restriction is suspicious if there are 8 underlying vowels, but makes sense if high vowels are surface allophones of non-high vowels, triggered by the presence of a voiced or aspirated consonant.

While there are some additional complications not treated here (see Cohn, 1993a; Davies, 2010; Kiliaan, 1897; Misnadin, 2016; Stevens, 1968 for more extensive discussion and examples), an analysis which permits the /D/ and /TH/-series plosives to function together as a distinct pair has clear advantages. However, this raises the question of what feature(s) these plosive series might share, since *a priori*, we would expect phonological rules to involve natural classes (Cohn, 1993b). Researchers have suggested that both types of plosive could involve a lowered larynx (Cohn, 1993a,b) and/or an advanced tongue root (Trigo, 1991), both of which would predict a range of acoustic effects including pitch lowering, vowel raising, and/or lax/breathy voice quality (Brunelle, 2010; Laver, 1980). In this respect, Madurese would resemble a 'register' system, common among languages of mainland Southeast Asia (Cohn and Lockwood, 1994; Henderson, 1952), in which some combination of pitch, voice quality, vowel quality, and durational differences are employed to distinguish (usually two) phonation types (Table III).

Table III. Typical acoustic correlates of register systems (after Brunelle and Kirby, 2016).

High register (voiceless plosives, *pa)	*Low register* (voiced plosives, *ba)
Shorter VOT	Longer VOT
Higher pitch	Lower pitch
Monophthongs/shorter vowels	Diphthongs/longer vowels
Raised F1/[-ATR]	Lowered F1/[+ATR]
Tense/modal voice	Lax/breathy voice

Previous acoustic descriptions (Cohn, 1993a; Cohn and Lockwood, 1994) concluded that Madurese bears the acoustic hallmarks of a register system. However, these findings were based on the speech of just one or two speakers, and in some cases run counter to phonetic expectations. For instance, Cohn and Lockwood (1994) report high onset $F0$ (C$F0$) following voiced stops (*contra* House and Fairbanks, 1953 and much subsequent work) as well as a reversed intrinsic $F0$ (I$F0$) effect, with high vowels supposedly having lower $F0$ than non-high vowels (*contra* Whalen and Levitt, 1995). If these findings are accurate, Madurese would be highly unusual. Moreover, if it is indeed a register system of the Southeast Asian type, it is especially interesting as in canonical register systems, onset differences in terms of voicing lead or lag are normally neutralized, with the contrastive function having shifted fully to spectral and/or temporal properties of the vowel (Huffman, 1976).

This paper presents a detailed study of the acoustic properties of Madurese obstruents and vowels, in order to better understand how the laryngeal contrast is realized in this language. In particular, we are interested if there is any acoustic evidence for an articulation shared by the /D/ and /TH/-series plosives in word-initial position. Our work builds on that of Cohn (1993a,b) and Cohn and Lockwood (1994), but uses a larger speaker sample and an expanded range of acoustic measures, giving special attention to dynamic measures of pitch, voice quality, and spectral properties of vowels.

B. Predictions

If Madurese /D/ and /TH/-series plosives share a common laryngeal configuration, such as a lowered larynx and/or advanced tongue root, they would be expected to share some, if perhaps not all of the 'low register' features shown in Table III. The articulatory mechanisms of tongue root advancement and larynx lowering are both predicted to produce similar acoustic consequences, including lowered F1, $F0$, and larger spectral balance differences (Denning, 1989; Guion *et al.*, 2004; Klatt and Klatt, 1990; Laver, 1980). Thus if /D/ and /TH/ share acoustic properties that are not simply expected due to the fact that both are followed by high vowels, we predict:

1. VOT will be longer for /TH/ than for /T/;

2. $F0$ will be lower following /D/ and /TH/ compared to /T/;

3. Vowels will be breathier following /D/ and /TH/ compared to /T/, as evidence by steeper spectral slopes;

4. Vowels will be longer after /D/ and /TH/ compared to /T/.

To anticipate our findings, the acoustic analyses revealed no single cluster of acoustic properties corresponding transparently to the phonological behavior of Madurese consonants. We conclude with a discussion of the origins of this system; whether its description as a language with a three-way laryngeal contrast is warranted; as well as the implications of our data for variation and universals of VOT more generally.

II. ACOUSTIC STUDY

A. Sound system

Madurese is typically analysed as having 27 consonants (Table I). While there is some debate about the precise place of articulation of some consonants, these differences do not concern us here; see Davies (2010); Misnadin and Kirby (2018) for discussion. All consonants can appear as word-medial geminates, but geminates never appear in word-initial position (Cohn and Ham, 1999), and so are not treated further here.

The eight surface vowel qualities [a ɛ ə ɔ ɤ i ɨ u] of Madurese can be organized into four pairs shown in Table II. Note that the pair [ə/ɨ] are significantly shorter than the others (see Sec. III E) and trigger obligatory gemination of a following consonant, possibly due to a syllable weight requirement (Misnadin and Kirby, 2018). For further details, see Cohn and Lockwood (1994); Davies (2010); Misnadin and Kirby (2018) and references therein.

B. Participants

Fifteen native speakers of Madurese from across four regencies in Madura (Bangkalan, Sampang, Pamekasan and Sumenep) were recorded for the study. They consisted of 8 females (mean age 20, range 18-21) and 7 males (mean age 22, range 20-28). All were undergraduate students at Trunojoyo University in Madura at the time of recording. None of the participants reported a history of hearing and speech disorders. They were paid for their effort and participation in the study.

Like nearly all Madurese speakers, the participants were also speakers of Standard Indonesian in formal settings such as in school and in other activities that involve speakers of different local languages. In addition, they also spoke some English at school and university. However, all participants grew up in dominantly Madurese-speaking households and mostly used Madurese in their daily lives. Although there is some variation between Madurese dialects, this is largely lexical and morphological in nature (Davies, 2010; Kiliaan, 1897; Soegianto *et al.*, 1986; Sutoko *et al.*, 1998); we know of no dialect differences that might impact the realization of the laryngeal contrast (though this is not to say that none exist).

C. Speech materials

188 Madurese words were selected for recording (see Supplementary Materials). The selection of words was done in such a way that voicing type, place of articulation and vowel type had comparable and adequate representations. We do not analyze any of the retroflex stops /ʈ ʈ$^{\mathrm{h}}$ ɖ/ because we were not able to find a representative sample of items with these plosives in absolute-initial position (/ɖ/ is especially rare).

All words are disyllabic with the syllable patterns of $C_1V_1C_2V_2$ and $C_1V_1C_2V_2C_3$ except *dupolo* 'twenty', which has three syllables, due to the difficulty of finding more words with similar place and vowel categories. Although differences in syllable type may affect vowel duration, this should not impact the consistency of the measurement results, as only the first syllable was analyzed. Where possible, we tried to insure that plosives in C_2 position were balanced in terms of place and voicing categories, in order to minimize any effects on the vowel of interest.

Target items were embedded in a sentence frame *Ngèrèng maos* ___ *sè saè* [ŋɛrɛŋ maɔs ___ sɛ saɛ] 'Let's read ___ well'. They were presented in orthographic form using a presentation script that was set up to randomise them in three blocks. Participants were instructed to read the sentences as fluently and naturally as possible. Recordings were made in a quiet room using a Marantz PMD661 portable audio recorder with a Shure SM10A head-mounted microphone and made in mono at a sampling rate of 44,100 Hz with 16-bit resolution. In total, 8,460 tokens (15 speakers x 188 items x 3 repetitions) were targeted for recording. Due to some participants occasionally skipping an item in the script, 8,397 tokens were ultimately recorded and analyzed.

D. Acoustic measurements and analysis

For each token, the duration of C_1 and V_1, along with the point of voice onset, were hand measured based on the acoustic waveform. Parameter extraction was done for each participant using the PraatSauce suite (Kirby, 2018b). Pitch was estimated using Praat's autocorrelation method in the range 75 to 300 Hz. Formant resonances were estimated by the Burg LPC algorithm using a 10-pole filter and a Gaussian-like analysis window with an effective range of 25 ms. We used a formant ceiling of 5000 Hz for males and 5500 Hz for females, with bandwidths estimated using the formula of Hawks and Miller (1995).

As the production of breathy voice has been observed to attenuate low-frequency spectral components and boost high-frequency components (compared to modally phonated signals), we measured several harmonic amplitude components from the low-, mid-, and high-frequency regions of the signal (H1, H2, A1, A2, A3, H2k, H5k). Components were identified automatically using a peak-finding algorithm based on the long-term average spectrum calculated over a 25 ms window at each measurement point. We corrected the raw amplitudes of these components using the formula of Iseli *et al.* (2007); these are reported as H1*, H2*, etc. We also calculated the cepstral peak prominence (Hillenbrand *et al.*, 1994), another acoustic measure which has been found to correlate with breathiness, using a lower quefrency of $1/300 \approx 0.0033$ sec, parabolic interpolation for peak amplitude detection, and Theil's robust line fit method. For an overview of these and other acoustic measures of voice quality, see Garellek (2019); Misnadin (2016).

All measurements were taken at 1 ms intervals across both the occlusion phase (for voiced plosives) and the post-release period (for all tokens) for each item; these measurements were then binned into 11 equally-spaced regions and averaged. Statistical analyses were performed in R (R Core Team, 2014) using the packages lme4 (Bates *et al.*, 2014) and emmeans (Lenth, 2018). Note that due to the CV co-occurrence restriction, it is not possible to include Vowel as a fully crossed factor in the models. Instead, we include a factor Vowel Pair with four levels (ə-ɨ, ɔ-u, a-ɤ, ɛ-i), which allows us to examine possible difference in vowel quality on dependent variables. For some comparisons, this is equivalent to just comparing vowel qualities, but this is not possible if comparing properties of the /T/ series plosives to either of the other two.

III. RESULTS

For ease of exposition, the main text focuses on informative visual displays. Full descriptive and inferential statistics may be found in the Supplementary Materials, and/or replicated by the reader using the data and R code available at `https://edin.ac/2GEJYan`.

A. Closure voicing and VOT

Fig. 1 displays the distribution of closure voicing duration (for /D/) and VOT (for /T/ and /TH/). VOT values for voiceless unaspirated and aspirated plosives are seen to overlap quite extensively, giving the appearance of a unimodal, if slightly skewed, distribution. This is a rather different pattern compared to most languages which are described as contrasting aspirated with unaspirated plosives, where the VOT ratio is normally on the order of 3 or 4:1 (Cho and Ladefoged, 1999; Kirby, 2018a; Lisker and Abramson, 1964). Distributions for both the voiceless aspirated and unaspirated series, which are often tightly clustered around a mean value in other languages with a three-way contrast, are well-fit by a gamma distribution (see Supplementary Materials).

About 9% (208/2322) of phonologically voiced plosives in the data were produced without any clear closure voicing. These are primarily instances of the palatal /ɟ/ (130 tokens, well over half of all such instances), which has mean and median of -58 ms with these tokens removed. Estimates for the other voiced plosives are also slightly longer (on the order of a few msec). There were no instances of /T/ or /TH/ coded as being produced with closure voicing, partial or otherwise.

As a further check on our annotations, we determined for each token the number of bins in the closure phase for which f0 was measurable. A very small number (2%) of voiceless tokens are found to occur with measurable periodicity during the closure, although closer inspection suggests many of these are spurious results reported by Praat's autocorrelation-based f0 tracker. There were virtually zero instances of voicing during the closure phase of aspirated plosives, consistent with the observations of Cohn and Lockwood (1994). Interestingly, closure voicing for voiced plosives is fairly evenly distributed, with roughly the same number of fully voiced closures as fully devoiced closures. Inspection of individual differences (see Supplementary Materials, Appendix B) shows that this is not uniform across speakers: a few participants (F5, M4, M6) have a greater proportion of devoiced than voiced /D/ closures, while for another (F4) the opposite trend is observed. For the remaining speakers, however, the distribution is more or less uniform.

To numerically assess the differences between the distributions, we fit a mixed model with factors Place (with levels Bilabial, Coronal, Palatal, Velar), Voice (with levels Voiced, Voiceless, Aspirated) and Vowel Pair (with levels ə-ɨ, ɔ-u, a-ɤ, ɛ-i) and all two- and three-way interactions, along with by-speaker slopes for Voice, Place, and Vowel Pair and by-item intercepts; this was the maximal model justified by the data. Averaging over Place, VOTs for /TH/ are consistently and significantly longer than /T/ by 15 to 25 ms. Averaging over Voice, the expected place-based asymmetries are observed: /p $\mathrm{p^h}$ t $\mathrm{t^h}$/ have shorter VOTs than /c $\mathrm{c^h}$ k $\mathrm{k^h}$/, respectively. For /D/, voicing lead is longest for bilabials, followed by velars, coronals and palatals; pairwise comparisons are all significantly different, but rather small (especially if devoiced tokens are disregarded). Notably, when averaging

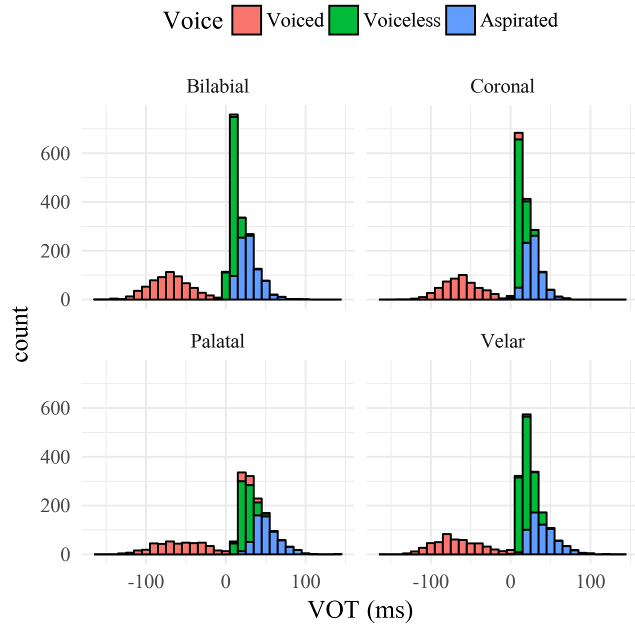

Figure 1. (Color online) Closure voicing duration/VOT of Madurese plosives by place of articulation and voicing type.

over Place, differences by Vowel Pair are minimal, and are significant primarily for aspirated plosives: VOT is longest when the following vowel is front [i] or back [u] (25-66 ms, depending on place of articulation) and around 10-15 ms shorter when preceding [ɨ] or [ɤ]. Estimated marginal means are provided in the Supplementary Materials (Appendix C).

B. Closure duration

Mean closure duration (Fig. 2) was significantly longer for /D/ at all places of articulation (from 7-32 ms on average). However, as described in Sec. III A, voicing was not always present for the entire closure. Voiceless bins were more common at the onset of closure, probably due to the preceding voiceless fricative in the carrier phrase (Fig. 3). For /D/,

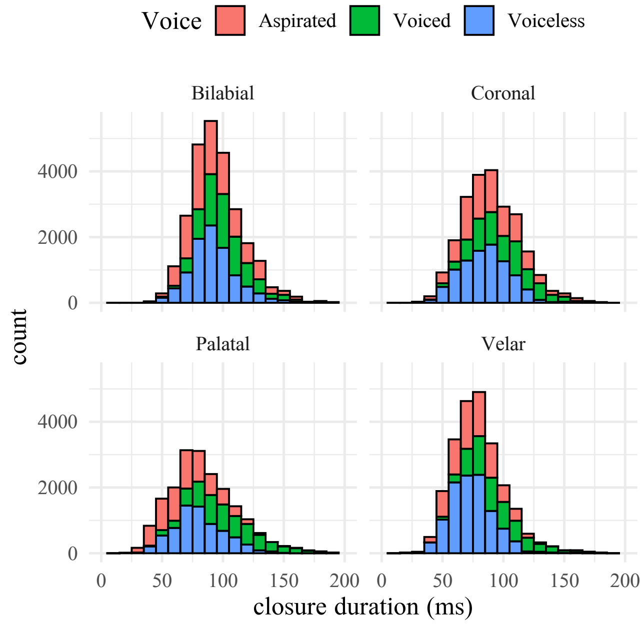

Figure 2. (Color online) Closure duration by place of articulation and voicing type.

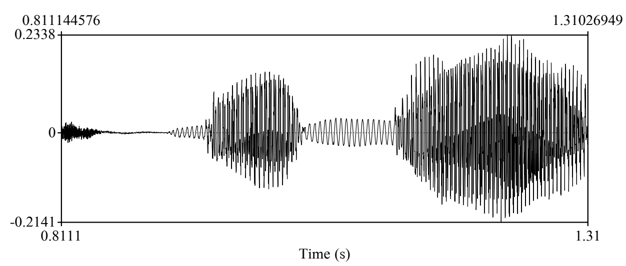

Figure 3. Example of token *bâbâ* [bɤbɤ] ‘under’, speaker F8. Frication from preceding sibilant fricative of carrier phrase shown at left edge.

there is a weak correlation between the number of voiced bins and closure duration (mean by-speaker $r^2 = 0.29$ with range 0.14 to 0.49) but a much stronger correlation between number of bins and actual duration of closure voicing (mean $r^2 = 0.75$, range 0.4 to 0.9) Durations for /T/ and /TH/ were usually indistinguishable, the exception being for palatals, where voiceless /c/ was usually longer than aspirated /c$^{\mathrm{h}}$/ by about 9 ms.

C. Fundamental frequency (CF0 and IF0)

Fig. 4 plots the $F0$ trajectory over the vowel for each speaker (in semitones, z-scored by speaker mean). We do not present an aggregate plot because, as can be seen in the figure, there is considerable individual variation which would be obscured by averaging. For all speakers, $F0$ is generally low or rising following /D/ and high or falling following /TH/. Note that this differs from Cohn and Lockwood (1994), who report $F0$ following voiced and aspirated plosives to be uniformly lower than that following voiceless unaspirated plosives, but is consistent with many other reports of C$F0$ behavior (Hanson, 2009; Hombert, 1978; House and Fairbanks, 1953; Kingston and Diehl, 1994; Kirby and Ladd, 2016; Silverman, 1986).

Conversely, the post-release effect of /T/ on $F0$ varies with speaker. For the majority of speakers, it patterns with /TH/ in raising $F0$, but for a few speakers (F4, F5, M1) it patterns with /D/. Although we do not have comparative data from sonorants, we expect that the post-release $F0$ trajectories of both /T/ and /D/ would not deviate significantly from a sonorant baseline for these speakers.[1]

To visualize I$F0$ effects, Fig. 5 plots $F0$ as a function of vowel pair by voicing, averaged across speakers, repetitions, and place of articulation. Cohn and Lockwood (1994) report that the non-high vowels [ɛ ɔ a ə] have higher $F0$ than the high vowels [i u ɤ ɨ], contrary to expectation (Whalen and Levitt, 1995). This is the case only if the data from vowels following voiced and aspirated plosives are conflated, however. As seen in Fig. 5, $F0$ is clearly controlled by onset type: within each Vowel Pair, the difference in mean I$F0$

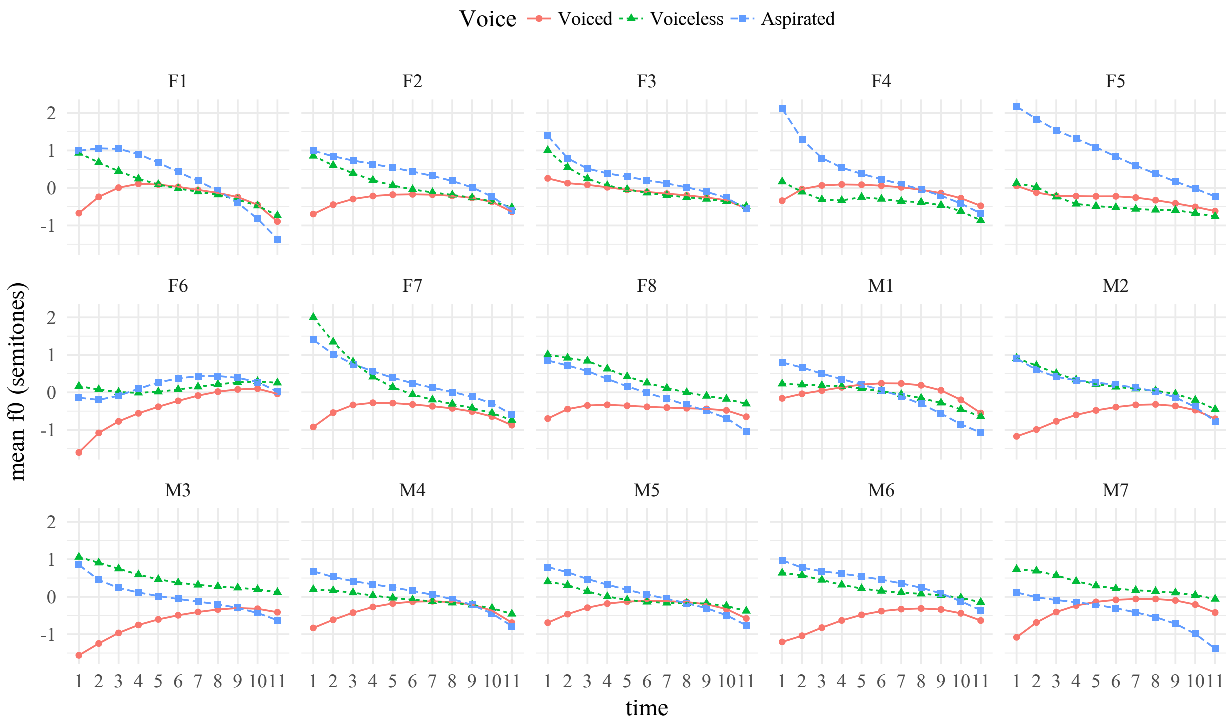

Figure 4. (Color online) $F0$ of Madurese plosives by place of articulation and voicing type, averaged over items and repetitions.

is not significantly different between voiceless and aspirated plosives (see Supplementary Materials for full model summaries). Once voicing type is controlled for, the expected I$F0$ effects more or less obtain. Notable is the behavior of the short mid vowel pair [ə/ɨ]: following voiceless plosives, estimated $F0$ is invariably quite high, while following voiced plosives it is generally lower.

D. Vowel quality

Fig. 6 shows the evolution of F1 and F2 over the V_1 vowel by voicing and vowel type, averaged over speakers, place of articulation, and repetitions. The pairs [a/ɤ], [ɛ/i] and [ɔ/u]

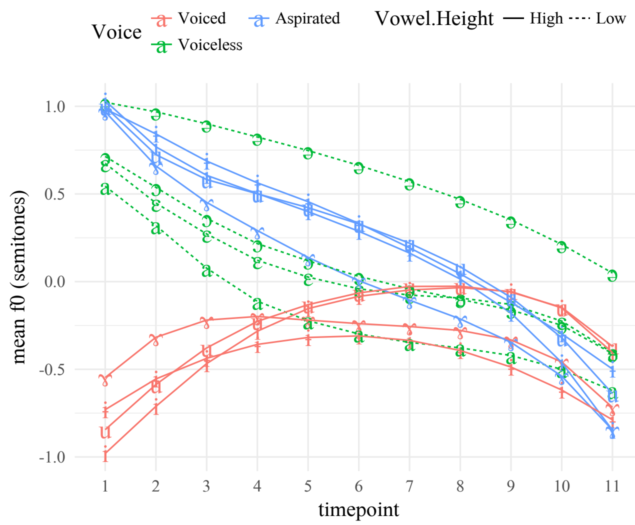

Figure 5. (Color online) I$F0$ by voicing and vowel, averaged over speakers, items, and repetitions.

are all clearly distinguished by F1: non-high [a], [ɛ], and [ɔ] all have predictably higher F1 values on the order of 200-300 Hz compared to [ɤ], [i], and [u], while [ə] has F1 of 125-130 Hz higher than [ɨ] (cf. Cohn, 1993b). The primary feature distinguishing [ə] from [ɨ] is F2, with [ɨ] having a more fronted realization (Misnadin and Kirby, 2018). Systematic F2 differences are also seen for [ɛ/i] and (to a lesser extent, and at voicing onset) for [a/ɤ], but not for [ɔ/u].

E. Vowel duration

The register interpretation predicts shorter vowels following high register (tense/voiceless) plosives and longer vowels following lower (lax/voiced) plosives. Fig. 7 shows the distribution of vowel length by voicing type. Vowels following voiced plosives are longest, followed by voiceless and then aspirated. Vowel length differences between voiced and aspirated plosives

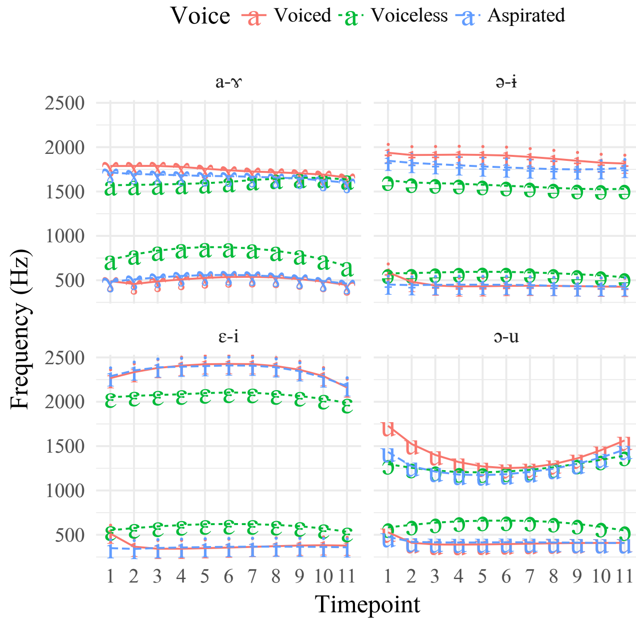

Figure 6. (Color online) F1 and F2 (in Hz) by voicing and vowel quality.

are on the order of 20 ms, except for the central pair [ɛ/i] which are always approximately half the duration of other vowels regardless of preceding plosive type.

F. Voice quality

We calculated eight measures of voice quality: H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, H2*-H4*, H2KHz-H5KHz, harmonics-to-noise ratio (HNR), and cepstral peak prominence (CPP). Exploratory data analysis (see Supplementary Materials, Appendix E) suggested that H1*-H2*, H2KHz-H5KHz, and CPP appeared to pattern together for the voiced and aspirated series. However, as shown in Fig. 8, this effect interacts with *phonetic* vowel height, not just vowel pair membership. For H1*-H2*, the high vowels [i ɨ u] have the highest amplitude differences, but the mid vowel [ɤ] patterns more closely with the other

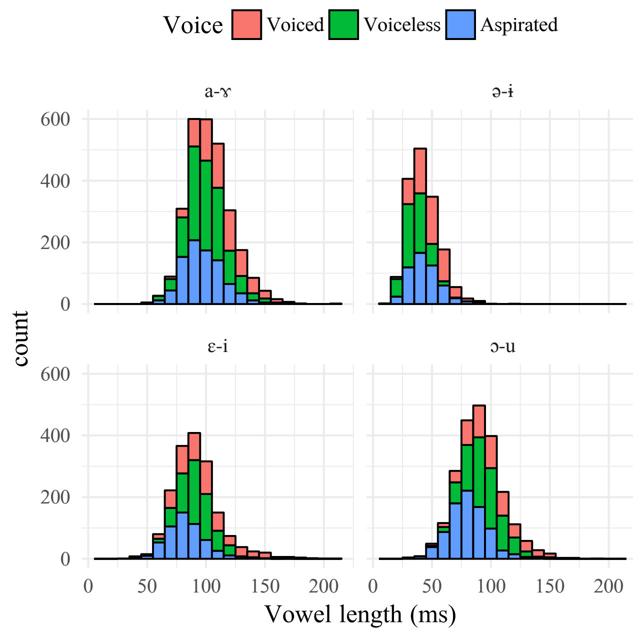

Figure 7. (Color online) Vowel duration (in ms) by voicing and vowel quality.

mid and low vowels. For H2KHz-H5KHz, large differences are observed between [ɛ] and [i], and slightly smaller, but still robust differences between [a] and [ɤ]; the more global patterning is one of [i u ɔ] vs. [a ɤ ə ɨ ɛ]. For CPP, differences are apparent primarily for [ɛ-i], and to a lesser extent [a/ɤ], but not for the central or back rounded vowel pairs. For CPP, [ɛ] and [ɔ] are distinct from [i] and [u] in the expected direction (the more prominent the cepstral peak, the stronger the harmonic content, so CPP should be lower for breathier vowels). However, no differences are apparent for the central vowel pairs.

IV. DISCUSSION

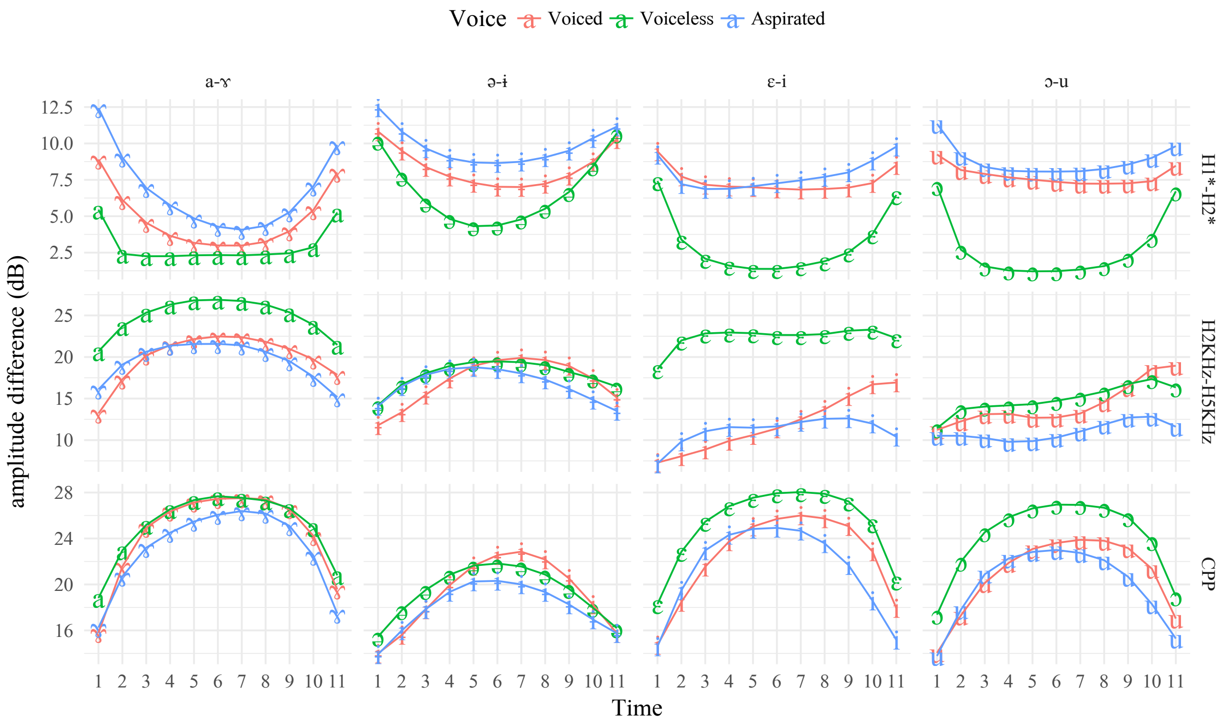

Figure 8. (Color online) Spectral measures of voice quality by voicing and vowel pair: H1*-H2*, H2KHz-H5KHz, CPP.

A. Summary of results

An overview of the findings is given in Table IV. /TH/ and /D/ pattern together in terms of vowel height and (for some vowel qualities) H1*-H2*, H2K-H5K and CPP, while /TH/ and /T/ pattern together in terms of $F0$ and closure duration. The VOT distributions for /T/ and /TH/, while statistically distinguishable, are heavily overlapping. We find no evidence that /TH/ plosives are realized with closure voicing, at least in word-initial, utterance-medial position. However, a small percentage of /D/-series plosives were sometimes devoiced in this context, probably due to the presence of a preceding voiceless fricative in the carrier phrase.

Table IV. Summary of acoustic findings by measure and phonation type.

	Onset		
Measure	/b d ɟ g/	/p t c k/	/p^h t^h c^h k^h/
VOT	-40–70 ms	10–25 ms	30–50 ms
Closure duration	95–105 ms	75-90 ms	70–95 ms
$F0$	Low	High[a]	High
H1*-H2*	High	Low	High
H2K-H5K	Low	High	Low
CPP	Lower	Higher	Lower
Vowel height	High	Low	High
Vowel duration[b]	Long	Shorter	Shortest

[a] For 12 of 15 speakers.

[b] Ignoring the short central vowel pair [ə/ɨ].

For most of the speakers in our sample, /p t c k/ and /p^h t^h c^h k^h/ appear to be realized similarly in terms of those properties unrelated to the height of the following vowels. In particular, these two series condition similar $F0$ contours, suggesting similar laryngeal tension settings, and similar VOTs, suggesting similar glottal aperture targets (on this see Section IV B below). For 3 of the 15 speakers, however, $F0$ for /T/ patterns with /D/, rather than with /TH/. The distinction between /T/ and /TH/ for these speakers is reminiscent of the tense/lax or stiff/slack distinction in Javanese (Fagan, 1988; Seyfarth *et al.*, 2017), but none of the speakers in our sample reported any fluency in this language.

Overall, our findings are largely consistent with those of Cohn and colleagues (Cohn, 1993a,b; Cohn and Ham, 1999; Cohn and Lockwood, 1994), with the important difference that our C$F0$ and I$F0$ results conform to the cross-linguistically expected patterns. An explanation for the I$F0$ differences was offered in Sec. III C, but what might account for the C$F0$ differences? For both speakers in Cohn and Lockwood (1994), C$F0$ for /b/ and /p$^{\mathrm{h}}$/ are 10-40 Hz lower at vowel onset compared to /p/ (and, unexpectedly, /m/). While it is possible that this represents regional variation, this seems unlikely given that the speakers in our sample come from across the island. However, as the data for our study was collected at 25 years after Cohn's recordings were made, generational differences cannot be ruled out (cf. Coetzee *et al.*, 2018). It is also possible that the differences between the carrier phrases in the two studies ("read X partway" vs. "let's read X again") may have altered the intonational context; and as previously noted, the immediate phonetic contexts are not identical (the preceding segment is vowel in Cohn's studies, and a voiceless fricative in ours). We hope to address these possibilities in future data collection.

Madurese does not appear to make a distinction in terms of voice quality that is independent from vowel quality. As shown in Sec. III F, those voice quality measures which do at first blush differentiate vowels following /D/ and /TH/ from /T/ are highly sensitive to vowel quality, primarily F1. Perhaps more tellingly, the fact that the differences are greatest during the steady-state portion of the vowel, rather than at the onset, further suggests they are driven by vowel quality, rather than by an articulation associated with the onset

(Blankenship, 2002; Garellek and Keating, 2011), which is what would be expected of a 'true' register language (Brunelle *et al.*, 2019).

B. Two or three plosives in Madurese?

Cho and Ladefoged (1999), surveying the distribution of VOT in 19 languages, conclude that only three modal phonetic categories of VOT are necessary – [voiced], [voiceless unaspirated], and [voiceless aspirated] – since no language makes contrastive use of more than two degrees of glottal aperture. At the same time, languages which do contrast the [unaspirated] and [aspirated] types typically choose modal values which are either well-separated in VOT space, such as Thai or English, or which recruit other acoustic dimensions to signal the contrast, such as Korean (Lisker and Abramson, 1964). Madurese appears to be a language more on the Korean model, in that it has recruited an orthogonal phonetic property (F1) to be the primary signal of contrast between two of its phonological categories. Do speakers then really maintain distinct glottal aperture targets for these two series?

We expect the answer is probably no, but then we are left needing to explain the stability of the VOT differences. At least three (non-mutually exclusive) factors could be involved:

1. Orthography. Aspiration is indicated in nearly all Madurese orthographies developed since the colonial period, although it was notably absent from the 1973 'standard' orthography (see Davies, 2010, 51–60).[2] Orthography can influence both speech production and word recognition (see Rastle *et al.*, 2011 for a recent review) and can potentially condition small but reliable differences in phonetic realization (Ernestus and Baayen, 2006; Warner *et al.*, 2006). The presence of an orthographic difference could thus help to maintain a phonetic contrast. That having been said, these sounds are orthographically represented as *voiced* aspirates, but we found no evidence that these sounds are realized with systematic closure voicing (cf. Sec. IV C below).

2. Vowel height differences. All else being equal, high, close vowels will offer greater aerodynamic resistance and could lead to a delay in the transglottal pressure drop necessary to initiate and sustain voicing (Ohala, 1981). This predicts VOT should be greater following high as opposed to low vowels. Correlations between vowel height and VOT have been documented for several languages including English (Klatt, 1975), French (Nearey and Rochet, 1994), and Hindi (Ohala and Ohala, 1992). In French, a language where voiceless stops are prototypically short-lag, Nearey and Rochet (1994) report mean differences of around 20 ms between the vowel pairs /i/ and /ɛ/ and /ɔ/ and /u/ following /p t k/, very similar to what we report in Sec. III A.[3] Berry and Moyle (2011) discuss how the mechanical relationship between vowel articulation and intrinsic $F0$ proposed by Honda (1983) might be extended to explain these effects: if contraction of the genioglossus and extrinsic laryngeal muscles increases vocal fold tension (and thereby phonation threshold pressure), this could in turn delay voicing onset, leading to longer VOTs before higher vowels.

3. Perceptual enhancement. A third possibility is that the VOT differences could be a listener-oriented enhancement (Diehl and Kluender, 1989; Kingston and Diehl, 1994): speakers lengthen the lag before high vowels to make the onset of the following vowel breathy, thereby increasing spectral tilt and enhancing the low frequency concentration of energy brought about by high vowels' low F1. This hypothesis makes what should be a testable perceptual prediction: differences in spectral tilt should condition similar shifts in listeners' categorization functions as do differences in voicing lag time.

Given these possibilities, we cautiously suggest that—for at least some speakers—Madurese specifies just a single glottal aperture target for both types of voiceless plosive. In models such as those proposed by Keating (1984) or Cho and Ladefoged (1999), this could be captured by a single context-restricted feature [voiceless]. The acoustic differences are then presumably the result of processes like those outlined above, i.e. effects of vowel height difference and/or perceptual enhancements. However, we also found evidence that /p t c k/ and /p^h t^h c^h k^h/ may involve complementary laryngeal settings: for three of the speakers in our study, /p t c k/ does not condition $F0$ raising in the following vowel, suggesting that these speakers may have distinct laryngeal tension targets for these categories.

All this raises the question of whether VOT is used by Madurese listeners in distinguishing between voiceless and aspirated plosives. In a pair of pilot experiments (Kirby and Misnadin, 2019), we found that Madurese listeners do not appear to attend to differences in positive VOT, even when vowel quality is ambiguous. This is consistent with a phonetic account on which the acoustic differences in VOT are the result of (language-specific or universal) physiological and aerodynamic processes.

However, we stress that, while the laryngeal contrast might be described as a two-way system phonetically (for at least some speakers), this is clearly inadequate from the phonological standpoint. We know of no evidence to suggest that the CV co-occurrence restriction is being systematically relaxed. This restriction is characteristic of some 95% of the Madurese lexicon (Stevens, 1968); the small number of exceptional items are mostly borrowings, and even some of these have alternants which conform to the general pattern (Davies, 2010, p. 36).[4] Morphophonological processes, such as that conditioned by the actor voice prefix described in Sec. I A, remain robust and productive to this day. Some means of formally distinguishing /T/ from /TH/ is therefore required, even if our acoustic data are not consistent with what might be expected of a phonetically grounded feature (e.g. [lowered larynx]).

C. Diachronic considerations

The historical source of the Madurese CV co-occurrence restriction remains debated. Comparative evidence suggests that Madurese items with /b/ are cognate with Javanese /w/, while Madurese /pʰ/ corresponds to Javanese /b/ (compare Javanese /wilaŋ/ ∼ Madurese [bitɔŋ] 'to count' but Javanese /bagus/ ∼ Madurese [pʰɤkʰus] 'good'). This led Stevens (1966) to posit two possibilities: either the common proto-language had two phonemes, *b (which became Javanese /w/ and Madurese /b/) and *B (which became Javanese /b/ and Madurese /pʰ/); or there was only *b, which became Javanese /w/ and Madurese /b/, with Madurese /pʰ/ introduced from subsequent borrowing of items with slack-voiced Javanese /b/. However, for Proto-Malayo-Polynesian *d and *g, the evidence points towards the aspirates as the Madurese reflexes, with instances modern /d/ and /g/—already comparatively relatively rare in Madurese, according to Kiliaan—as borrowings from Arabic and/or Malay (Kiliaan, 1897, p. 62 *ff.*; Stevens, 1966, p. 154).

Sorting out this complex state of affairs remains a challenge for the comparative Austronesianist, but we cautiously offer some speculation based on the present study. Regardless of the sources of the segments and the relative chronology of their introduction to the language, it seems Madurese must have at one time had a three-way phonetic contrast between (voiceless) fortis, (voiced) lenis, and something like breathy-voiced onsets. This would be consistent with the orthography developed in the colonial period, which represents these sounds as *bh, dh*, etc.[5] Subsequently, articulatory maneuvers to sustain voicing for both the latter series could have conditioned the perceptually (Lotto *et al.*, 1997) and typologically (Denning, 1989) expected changes in vowel height. Once the vowel height differences were phonologized, the redundant voicing for what is now the /TH/ series could be lost or variably realized (Brunelle *et al.*, 2019; Seyfarth *et al.*, 2017) (although recall that we did not find any evidence for variable realization in this data sample). The introduction of (something like) [b^{ɦ} d ɖ^{ɦ} g] to a system already containing [b d^{ɦ} ɖ g^{ɦ}] may have put pressure on the voiced aspirates to devoice, in order to enhance the contrast between items like *bhuta* [p^{h}uta] 'giant' and *buta* [buta] 'blind' (which on this account would have once been something like [b^{ɦ}ɔta] and [bɔta], respectively). The voiced series might plausibly have resisted devoicing if there was prestige associated with accurate pronunciation of borrowed items (cf. the history of non-allophonic /v/ in English). In effect, the voiced aspirates would have merged with the voiceless unaspirates, with the modern VOT differences persisting for aerodynamic reasons (Sec. IV B).[6] Seen in this way, the synchronically unusual CV co-occurrence restriction may be understood as a having arisen through the stepwise phonologization of common phonetic effects (see e.g. Bach and Harms, 1972; Blevins, 2004; Hyman, 2001; Jacques, 2013; Yu, 2004 and references therein).

V. SUMMARY

We find no evidence that the voiced and voiceless aspirated plosives of Madurese condition a unique constellation of acoustic features, beyond the fact that both participate in the same phonotactic pattern with respect to vowel height. The acoustic properties they do have in common—limited to a few measures of spectral balance–are most likely an artifact of the fact that they are always followed by the same subset of high, close vowels. Thus, it is unlikely that these segments are synchronically characterized by a common articulatory gesture, such as a lowered larynx or advanced tongue root, although it is possible that they shared such an articulation at some point in the past.

In terms of VOT, closure duration, and $F0$ effects on the following vowel, on the other hand, Madurese voiceless aspirated and unaspirated plosives are acoustically rather similar. Thus, phonetically speaking, Madurese can be described as contrasting prevoiced with voiceless plosives, but two types of 'voiceless plosive' must be distinguished phonologically. Diachronically, this state of affairs most likely developed as a kind of register system, albeit one which was heavily influenced by borrowing at a critical stage in its evolution.

[1]This is based on the assumption that sonorants are the segments least likely to perturb f0 away from its intonationally specified baseline, because the lack of complete supraglottal occlusion would not require any laryngeal adjustments designed to increase the volume of the supraglottal cavity for the purposes of ensuring a transglottal pressure differential suitable to sustain vocal fold vibration. This is predicated on the assumption the CF0 effects may be caused by changes in vocal fold tension (e.g. Löfqvist *et al.*, 1989). Moreover, as the nasal cavity offers little resistance to airflow, nasals are not expected to exert

significant change on oral air pressure which, due to decreasing the transglottal pressure differential, has been hypothesized to perturb pitch via aerodynamic means (Ohala, 1975).

[2]The Madurese orthography used in this article is the version ratified at the 2008 *Kongres Bahasa Madura Internasional*. This orthography distinguishes all three plosives types, but does not have separate graphemes for [ə] and [ɨ].

[3]This generalization does not hold for the pair /pi/-pɛ/ in Nearey and Rochet (1994)'s data, but this may be an outlier; cf. Fischer-Jørgensen (1972).

[4]In connected speech, apparent height harmony violations may also be introduced by the coarticulatory influence of an adjacent palatal glide; see Misnadin and Kirby (2018).

[5]Note that the CV co-occurrence restriction was clearly established well before the colonial period, as the orthography also indicates the vowel height differences. Kiliaan (1897, pp. 2-3) describes /D/ and /TH/ as *zachte klemletters* distinguished by presence vs. absence of aspiration; whether *zacht* should be interpreted as 'voiced' or simply something like 'lenis' is unclear.

[6]Pittayaporn and Kirby (2017) document just such a shift for a Tai language of Vietnam, in which the historical breathy voiced onsets appear to have lost their voicing and merged with the voiceless unaspirated series (albeit without a concomitant shift in vowel quality).

Adisasmito-Smith, N. (**2004**). "Phonetic and phonological influences of Javanese on Indonesian," Ph.D. thesis, Cornell University.

Bach, E., and Harms, R. T. (**1972**). "How do languages get crazy rules?," in *Linguistic Change and Generative Theory*, edited by R. P. Stockwell and R. K. S. Macaulay (Indiana University Press, Bloomington, IN), pp. 1–21.

Bates, D., Maechler, M., and Bolker, B. (**2014**). *lme4: Linear mixed-effects models using Eigen and S4.*, http://www.CRAN.R-project.org/package=lme4.

Berry, J., and Moyle, M. (**2011**). "Covariation among vowel height effects on acoustic measures," Journal of the Acoustical Society of America **130**(5), EL365–EL371, doi: 10.1121/1.3651095.

Blankenship, B. (**2002**). "The timing of nonmodal phonation in vowels," Journal of Phonetics **30**, 163–191, doi: 10.006/jpho.2001.0155.

Blevins, J. (**2004**). *Evolutionary phonology* (Cambridge University Press, Cambridge).

Brunelle, M. (**2010**). "The role of larynx height in the Javanese tense - lax stop contrast," in *Austronesian and theoretical linguistics*, edited by R. Mercado, E. Potsdam, and L. deMena Travis (John Benjamins), pp. 7–23.

Brunelle, M., and Kirby, J. (**2016**). "Tone and phonation in Southeast Asian languages," Language and Linguistics Compass **10**(4), 191–207, doi: 10.1111/lnc3.12182.

Brunelle, M., Tấn, T. T., Kirby, J., and Đinh Lư Giang (**2019**). "Obstruent devoicing and registrogenesis in Chru," in *Proceedings of the 19th International Congress of Phonetic Sciences*.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: evidence from 18 languages," Journal of Phonetics **27**(2), 207–229, doi: 10.1006/jpho.1999.0094.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., and Wissing, D. (**2018**). "Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis," Journal of Phonetics **66**, 185–216, doi: 10.1016/j.wocn.2017.09.009.

Cohn, A. C. (**1993**a). “Consonant-vowel interactions in Madurese: the feature lowered larynx,” in *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*, edited by K. Beals (Chicago Linguistic Society), pp. 105–119.

Cohn, A. C. (**1993**b). “Voicing and vowel height in madurese: A preliminary report,” in *Oceanic Linguistics Special Publications, No. 24: Tonality in Austronesian Languages*, edited by J. A. Edmondson and K. J. Gregerson (University of Hawaii Press), pp. 107–121.

Cohn, A. C., and Ham, W. H. (**1999**). “Temporal properties of Madurese consonants: a preliminary report,” in *Selected Papers from the Eight International Conference on Austronesian Linguistics*, edited by E. Zeitoun and P. J.-K. Li, Academica Sinica, Vol. 1, pp. 227–249.

Cohn, A. C., and Lockwood, K. (**1994**). “A phonetic description of Madurese and its phonological implications,” Working Papers of the Cornell Phonetics Laboratory **9**, 67–92.

Davies, W. D. (**2010**). *A grammar of Madurese* (De Gruyter, Berlin).

Denning, K. (**1989**). “The diachronic development of phonological voice,” Ph.D. thesis, Stanford University.

Diehl, R. L., and Kluender, R. (**1989**). “On the objects of speech perception,” Ecological Psychology **1**, 121–144.

Ernestus, M., and Baayen, R. H. (**2006**). “The functionality of incomplete neutralization in Dutch: The case of past-tense formation,” in *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. Goldstein, D. H. Whalen, and C. T. Best, Phonology and Phonetics (Mouton de Gruyter), pp. 27–49.

Fagan, J. L. (**1988**). "Javanese intervocalic stop phonemes: The light/heavy distinction," in *Studies in Austronesian Linguistics*, edited by R. McGinn (Ohio University Center for International Studies), pp. 173–202.

Fischer-Jørgensen, E. (**1972**). "p t k et b d g français en position vocalique accentuée," in *Papers in linguistics and phonetics to the memory of Pierre Delattre*, edited by V. Albert, Janua Linguarum (Mouton), pp. 143–200.

Garellek, M. (**2019**). "The phonetics of voice," in *The Routledge handbook of phonetics*, edited by W. F. Katz and P. F. Assmann (Routledge, New York), pp. 75–106.

Garellek, M., and Keating, P. (**2011**). "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec," Journal of the International Phonetic Association **41**(2), 185–205, doi: 10.1017/S0025100311000193.

Guion, S. G., Post, M. W., and Payne, D. L. (**2004**). "Phonetic correlates of tongue root vowel contrasts in Maa," Journal of Phonetics **32**(4), 517–542, doi: 10.1016/j.wocn.2004.04.002.

Hanson, H. M. (**2009**). "Effects of obstruent consonants on fundamental frequency at vowel onset in English," Journal of the Acoustical Society of America **125**(1), 425–441, doi: 10.1121/1.3021306.

Hawks, J. W., and Miller, J. D. (**1995**). "A formant bandwidth estimation procedure for vowel synthesis," Journal of the Acoustical Society of America **97**(2), 1343–1344.

Henderson, E. J. A. (**1952**). "The main features of Cambodian pronunciation," Bulletin of the School of Oriental and African Studies, University of London **14**(1), 149–174.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," Journal of Speech, Language, and Hearing Research **37**(4), 769–778.

Hombert, J.-M. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic Press, New York), pp. 77–111.

Honda, K. (**1983**). "Relationship between pitch control and vowel articulation" Haskins Laboratories Status Report on Speech Research SR-73, 269–282.

House, A. S., and Fairbanks, G. (**1953**). "The influence of consonant environment on secondary acoustical characteristics of vowels," Journal of the Acoustical Society of America **25**, 105–113.

Huffman, F. E. (**1976**). "The register problem in fifteen Mon-Khmer languages," in *Austroasiatic studies, part I*, edited by P. N. Jenner, L. C. Thompson, and S. Starosta (The University Press of Hawaii, Honolulu), pp. 575–589.

Hyman, L. (**2001**). "The limits of phonetic determinism in phonology: *NC revisited," in *The Role of Speech Perception in Phonology*, edited by E. Hume and K. Johnson (Academic Press, San Diego), pp. 141–185.

Iseli, M., Shue, Y.-L., and Alwan, A. (**2007**). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," Journal of the Acoustical Society of America **121**(4), 2283–2295.

Jacques, G. (**2013**). "The sound change *s- ¿ n- in Arapaho," Folia Linguistica Historica **34**(1), 43–57, doi: 10.1515/flih.2013.002.

Keating, P. A. (**1984**). "Phonetic and phonological representation of stop consonant voicing," Language **60**(2), 286–319.

Kiliaan, H. N. (**1897**). *Madoereesche spraakkunst. Eerste stuk: inleiding en klankleer* (Landsdrukkerij, Batavia).

Kingston, J., and Diehl, R. L. (**1994**). “Phonetic knowledge,” Language **70**, 419–454.

Kirby, J. (**2018**a). “Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages,” Journal of Phonetics **71**, 326–354, doi: 10.1016/j.wocn.2018.09.009.

Kirby, J. (**2018**b). “Praatsauce: Praat-based tools for spectral analysis [computer program]” Retrieved 6 April, 2019 from https://github.com/kirbyj/praatsauce.

Kirby, J., and Ladd, D. R. (**2016**). “Effects of obstruent voicing on vowel f0: evidence from ‘true voicing’ languages,” Journal of the Acoustical Society of America **140**(4), 2400–2411.

Kirby, J., and Misnadin (**2019**). “Perception of laryngeal contrast in Madurese,” in *Proceedings of the 19th International Congress of the Phonetic Sciences*, Melbourne.

Klatt, D. H. (**1975**). “Voice onset time, frication, and aspiration in word-initial consonant clusters,” Journal of Speech, Language, and Hearing Research **18**(4), 686–706.

Klatt, D. H., and Klatt, L. C. (**1990**). “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” Journal of the Acoustical Society of America **87**(2), 820–857.

Kulikov, V. (**2010**). “Voicing and vowel raising in Sundanese,” in *Proceedings of the 17th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA XVII).*

Laver, J. (**1980**). *The Phonetic Description of Voice Quality* (Cambridge, Cambridge University Press).

Lenth, R. (**2018**). *emmeans: Estimated Marginal Means, aka Least-Squares Means*, https://CRAN.R-project.org/package=emmeans, R package version 1.1.2.

Lisker, L., and Abramson, A. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Löfqvist, A., Baer, T., McGarr, N., and Story, R. S. (**1989**). "The cricothyroid muscle in voicing control," Journal of the Acoustical Society of America **85**, 1314–1321.

Lotto, A. J., Holt, L. L., and Kluender, K. R. (**1997**). "Effect of voice quality on perceived height of English vowels," Phonetica **54**, 76–93.

Misnadin (**2016**). "The phonetics and phonology of the three-way laryngeal contrast in Madurese," Ph.D. thesis, University of Edinburgh.

Misnadin, and Kirby, J. (**2018**). "Madurese," Journal of the International Phonetic Association doi: 10.1017/S0025100318000257.

Nearey, T. M., and Rochet, B. L. (**1994**). "Effects of place of articulation and vowel context on vot production and perception for french and english stops," Journal of the International Phonetic Association **24**(1), 1–18, doi: 10.1017/S0025100300004965.

Ohala, J. J. (**1975**). *Phonetic explanations for nasal sound patterns*, 289–316 (Language Universals Project).

Ohala, J. J. (**1981**). "Articulatory constraints on the cognitive representation of speech," in *The cognitive representation of speech*, edited by T. Myers, J. Laver, and J. Anderson (North Holland), pp. 111–122.

Ohala, M., and Ohala, J. J. (**1992**). "Phonetic universals and Hindi segment duration," in *International Conference on Spoken Language Processing*, pp. 831–834.

Pawitra, A. (**2009**). *Kamus standard bahasa Madura-Indonesia* (Dian Rakyat).

Pittayaporn, P., and Kirby, J. (**2017**). "Laryngeal contrasts in the Tai dialect of Cao Bằng," Journal of the International Phonetic Association **47**(1), 65–85, doi: 10.1017/S0025100316000293.

R Core Team (**2014**). *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing), http://www.R-project.org/.

Rastle, K., McCormick, S. F., Bayliss, L., and Davis, C. J. (**2011**). "Orthography influences the perception and production of speech," Journal of Experimental Psychology: Learning, Memory, and Cognition **37**(6), 1588–1594, doi: 10.1037/a0024833.

Seyfarth, S., and Garellek, M. (**2018**). "Plosive voicing acoustics and voice quality in Yerevan Armenian," Journal of Phonetics **71**, 425–450, doi: 10.1016/j.wocn.2018.09.001.

Seyfarth, S., Vander Klok, J., and Garellek, M. (**2017**). "Acoustics of the tense-lax stop contrast in Semarang Javanese," Journal of the Acoustical Society of America **142**(4), 2581, doi: 10.1121/1.5014441.

Silverman, K. E. A. (**1986**). "F0 segmental cues depend on intonation: the case of the rise after voiced stops," Phonetica **43**, 76–91.

Soegianto, Soetoko, Soekarto, and Soetarto, A. (**1986**). *Pemetaan bahasa Madura di pulau Madura* (Pusat Pembinaan dan Pengembangan Bahasa, Jakarta).

Stevens, A. M. (**1966**). "The Madurese reflexes of Proto-Malayopolynesian," Journal of the American Oriental Society **86**(2), 147–156, doi: 10.2307/596427.

Stevens, A. M. (**1968**). *Madurese phonology and morphology* (American Oriental Society, New Haven).

Sutoko, Soegianto, Surani, S., Sariono, A., and Suyanto, B. (**1998**). *Geografi dialek bahasa Madura* (Pusat Pembinaan dan Pengembangan Bahasa, Jakarta).

Thurgood, E. (**2004**). "Phonation types in Javanese," Oceanic Linguistics **43**(2), 277–295, doi: 10.1353/ol.2005.0013.

Trigo, L. (**1991**). "On pharynx-larynx interactions," Phonology **8**(1), 113–136, doi: 10.1017/S0952675700001299.

Warner, N., Good, E., Jongman, A., and Sereno, J. (**2006**). "Orthographic vs. morphological incomplete neutralization effects," Journal of Phonetics **34**, 285–293.

Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic F0 of vowels," Journal of Phonetics **23**, 349–366.

Yu, A. C. L. (**2004**). "Explaining final obstruent voicing in Lezgian: Phonetics and history," Language **80**(1), 73–97.